

УДК 004.827

**ИССЛЕДОВАНИЕ И ОБРАБОТКА ЗАВИСИМЫХ СОБЫТИЙ И ВРЕМЕННЫХ
КОНСТРУКЦИЙ В МЕДИЦИНСКИХ ТЕКСТАХ ДЛЯ ФОРМИРОВАНИЯ
ВРЕМЕННОЙ ШКАЛЫ ХОДА ЛЕЧЕНИЯ ПАЦИЕНТА**

Исаков Т.Т. (Университет ИТМО, г. Санкт-Петербург)

Научный руководитель – к.т.н., Ковальчук С.В.

(Университет ИТМО, г. Санкт-Петербург)

Аннотация

Для построения временной шкалы хода лечения пациента необходимо уметь качественно извлекать временные конструкции и соответствующие им события. Одной из множества проблем является существование зависимых событий, порождающих неопределенность. В работе представлен один из возможных методов решения данной проблемы.

Введение. Обработка неструктурированного текста на данный момент все еще является трудновыполнимой задачей, требующей множества исследований и использования ансамблей различных методов обработки текста. Вопрос стоит особо остро, когда речь идет о специализированных текстах, таких как медицинские записи. В медицинских записях особое место занимают временные конструкции, указывающие на те или иные события в жизни пациента. Существующие на данный момент методы аннотации медицинских событий предусматривают работу с данными в однозначном, нормализованном виде. Однако, в анамнезах присутствует множество неопределенностей, решение которых требует разработки новых подходов.

Основная часть. Для построения временной шкалы необходимо иметь представление о том какие именно события встречаются в анамнезах. С помощью алгоритма поиска временных конструкций и соответствующих им событий был получен набор событий из анамнезов, который был представлен в векторном виде с целью построения графа близости. На основе графа близости для дальнейшей классификации были выделены основные классы событий. Существуют разные виды зависимостей многие из которых на данном этапе разрешить не представляется возможным. В данной работе рассматриваются те зависимые события, в которых присутствует ссылка на другое событие. Для таких записей используется алгоритм, основанный на правилах, в которых проверяется наличие соответствующего класса события среди всех событий в рамках записи. Точность определяется долей верно обработанных зависимостей для тестового набора. На данном этапе при имеющемся разнообразии повествования событий использование правил является наиболее верным методом разрешения зависимостей такого типа. Однако в дальнейшем не исключается использование методов на основе машинного обучения.

Выводы. Построение логически взаимосвязанной временной шкалы на основе медицинских записей является одной из важных задач извлечения информации из медицинских записей, так как построение такой шкалы может облегчить чтение записей медицинскими работниками. К тому же такие данные могут использоваться в дальнейших исследованиях для выявления новых взаимосвязей между событиями и обнаружения паттернов. Данные результаты являются основной для дальнейшей разработки алгоритмов разрешения зависимостей других видов, а также разрешения коллизий.

Исаков Т.Т. (автор)

Подпись

Ковальчук С.В. (научный руководитель)

Подпись

